4A

УДК 004.89:004.93

В.Г. Абакумов, Е.Ю. Ломакина
Национальный технический университет «КПИ», г. Киев, Украина

Автоматическое распознавание жестов в интеллектуальных системах

В работе рассмотрена концепция использования жестов рук в интеллектуальных системах, в частности для управления робототехническими устройствами. Проанализированы основные процедуры распознавания движений рук, выделены основные моменты на каждом из этапов распознавания. Предложен подход, позволяющий уменьшить размерность пространства, в котором находится рука, а также избежать ограничений, связанных с использованием специальных маркировочных средств.

Введение

В последнее время разработчики интеллектуальных систем все больше внимания уделяют автоматическому распознаванию жестов с помощью визуальных систем [1]. Такой интерес вызван природным характером и удобством использования интерфейса на основе жестов, а также возможностью его применения в большинстве областей человеческой деятельности. Постановка задачи распознавания жестов комплексная и учитывает неоднозначную природу статических и динамических жестов, проблемы выделения руки на окружающем фоне, условия освещения и помехи. Решение задачи в большинстве случаев предполагает выбор алгоритмов с использованием компьютерных ресурсов.

В статье рассмотрена проблема управления робототехникой с помощью заданного набора жестов. Поскольку управление происходит в режиме реального времени, то необходим алгоритм, не требующий огромных вычислительных затрат.

Ранние технологии распознавания жестов предполагали использование маркеров, прикрепленных к кончикам пальцев пользователя. С помощью соответствующего алгоритма определялись наличие маркера и его цвет, выполнялась идентификация пальцев, задействованных для формирования жеста [2]. Однако использование маркеров накладывает некоторые ограничения на работу пользователя, и, как следствие, преимущество стали предоставлять бесконтактным технологиям.

Современные методы используют более прогрессивную технику на основе компьютерного зрения. Распознавание жестов может осуществляться методом создания пространства кривых, суть которого заключается в нахождении граничных контуров руки [3]. Данный подход достаточно надежен и инвариантен к перемещениям и вращениям руки, однако требует больших вычислительных затрат. Также был предложен алгоритм распознавания положения руки с помощью изображений скелета руки [4]. В данном случае применяется многокамерная система для нахождения центра гравитации руки и наиболее отдаленных от него точек, обеспечивая, таким образом, информацию о положении кончиков пальцев, которая используется для построения изображений скелета руки и, соответственно, распознавания. Остальные методы распознавания жестов используют специальные методы сравнения [5], дескрипторы Фурье, нейросети, гистограммы положения [6], фильтрацию точек [7].

При распознавании используется фиксированный набор жестов, с помощью которого задаются определенные команды для управления робототехникой в режиме реального

времени. Поэтому быстродействие и простота алгоритма имеют большое значение. Такой подход включает сегментацию изображения руки на основе цветовых характеристик кожи и ограничений размерности [8], [9]. Ограничение размерности является необходимым условием, так как большая размерность вызывает огромные вычислительные затраты. Все движения руки описываются с помощью базисных векторов. Для уменьшения размерности пространства наблюдаемых векторов без существенной потери информации рекомендуется применять анализ главных компонент (АГК), а для отображения характерных особенностей – анализ независимых компонент (АНК) [9].

Для того чтобы система отвечала нужными реакциями на определенные совокупности внешних воздействий, необходимо подключить процесс обучения, который заключается в адаптации системы к конкретным движениям рук пользователя (заданному набору жестов). В качестве объектов обучения выступают визуальные изображения рук.

Итак, на основе этих предварительных процедур обработки генерируется сигнал, который несет информацию о жесте на изображении. Далее жест сравнивается с набором жестов из базы данных и, в случае успешной классификации, ему присваивается определенная команда. На выходе системы формируется управляющий сигнал, предающий команду, на основе которой автоматическое устройство выполняет то или иное действие.

Распознавание жестов рук

Предложенный алгоритм распознавания жестов состоит из следующих этапов:

1) обучение системы, задание конкретного набора жестов и соответствующих команд;

2) преобразование исходного изображения в начальное представление (математические преобразования, вычисление главных компонент);

3) локализация и сегментация областей руки на изображении на основе ключевых характеристик;

4) механизм классификации (моделирования).

1 Обучение системы

Рассмотрим задачу обучения и управления робототехникой. Автоматизированное устройство содержит видеокамеру, с помощью которой фиксируется положение руки пользователя. Алгоритм должен позволять роботу идентифицировать жест во входном изображении как одну из определенных команд. Каждая идентифицированная команда будет использоваться для управления роботом, выполнения тех или иных задач. Жестам могут быть присвоены различные значения в зависимости от функций робота. При задании команд также можно использовать информацию о количестве пальцев. Например, один палец может означать – «движение вперед», два – «назад», три – «направо», четыре – «налево», пять – «стоп».

Рекомендуется использование набора жестов из алфавита глухонемых, что позволит сделать работу с робототехникой, возможной для людей, не имеющих возможности пользоваться традиционными средствами интерфейса.

2 Преобразование исходного изображения

Для уменьшения размерности пространства наблюдаемых векторов без существенной потери информации применяется анализ главных компонент (АГК). Входные вектора представляют собой отцентрированные и приведённые к единому масштабу изображения рук. АГК состоит в линейном ортогональном преобразовании входного вектора X размерности N в выходной вектор Y размерности M, $N < M$. При этом компоненты вектора Y являются некоррелированными, а общая дисперсия после преобразования остаётся неизменной. Матрица X состоит из всех примеров изображений жестов обучающего набора.

Решив уравнение $\Lambda=\Phi^{T}\times\Sigma\times\Phi$, получаем матрицу собственных векторов Ф, где Σ – ковариационная матрица для X, а Λ – диагональная матрица собственных чисел. Выбрав из Ф подматрицу Φ_M, соответствующую M-наибольшим собственным числам, получим, что преобразование $y=\Phi_M^T\times\tilde{x}$, где $\tilde{x}=x-\overline{x}$ – нормализованный вектор с нулевым математическим ожиданием, характеризует большую часть общей дисперсии и отражает наиболее существенные изменения X.

Выбор первых M главных компонент разбивает векторное пространство на главное (собственное) пространство $F=\{\Phi_i\}_{i-1}^{M}$, содержащее главные компоненты, и его ортогональное дополнение $\overline{F}=\{\Phi_i\}_{i=M+1}^{N}$.

Входное изображение, с помощью вычисленных ранее матриц, разлагается на набор линейных коэффициентов, называемых главными компонентами. Сумма главных компонент, умноженных на соответствующие собственные вектора, является реконструкцией изображения.

Хотя анализ главных компонент эффективно используется для сокращения размерности пространства, с его помощью сложно отображать характерные особенности, так как его базисные векторы представляют глобальные характеристики. Для решения этой проблемы используется метод анализа независимых компонент (АНК).

Задачей анализа независимых компонент является разложение наблюдаемых случайных переменных x_j, описывающих движение рук, в линейную комбинацию независимых случайных величин u_k

$$x_j=\sum_{j=1}^{N}a_{ij}u_j=a_{i1}u_1+a_{i2}u_2+...+a_{iN}u_N. \tag{1}$$

Вместо сумм удобно использовать обозначения с вектор-матрицей, тогда уравнение (1) можно записать в виде:

$$x = A\times u,$$

где A – смешанная матрица с элементами a_{ij}; x – случайный вектор, компоненты которого – $x_1,\ldots,x_n$; u – случайный вектор с компонентами $u_1,..,u_n$.

Независимые компоненты являются неизвестными переменными. Также принято, что матрица смешивания неизвестна. С помощью случайного вектора x оценивается A и u.

Исходная точка для анализа независимых компонент – самое простое предположение, что компоненты u_j являются статистически независимыми. Также необходимо предположить, что независимый компонент должен иметь не Гауссово распределение. Тогда, после оценки матрицы A можно вычислить ее инверсную матрицу, обозначенную W, и получить независимый компонент:

$$u = W\times x.$$

Алгоритм вычисления независимых компонент опирается на центральную предельную теорему, утверждающую, что при определенных условиях сумма независимо распределенных случайных величин стремится к нормальному распределению по мере увеличения количества слагаемых. Используя это утверждение, поиск независимых компонентов, как линейных комбинаций наблюдаемых переменных, ведется таким способом, чтобы получить независимые случайные величины, распределение которых максимально далеко от нормального.

3 Сегментация изображения руки на основе ключевых характеристик

В качестве признака, который используется для отделения руки от фона на изображении, можно использовать цвет кожи. В данном случае для реализации сегментации применяется пиксельная модель кожи. Модель формируется исходя из информации

о цветности (тон и насыщенность), полученная в результате предварительного обучения, которое непосредственно состоит в размещении руки пользователя в области так называемого обучающего квадрата. Пиксели, заключенные в эту область, используются для обучения модели, после чего выделенные пиксели преобразовываются из цветового пространства RGB в пространство HSL, откуда затем получается информация о цветности.

Значения цветового тона H и насыщенности S для каждого выбранного пикселя образуют набор $\vec{x} = (\vec{x}_1, ..., \vec{x}_n)$, где n – количество отсчетов (пикселей), $\vec{x}_i = (H_i, S_i)$ – значения цветового тона и насыщенности i-го пикселя. Для представления функции плотности вероятности, описывающей принадлежность пикселей к цвету кожи, выбрана гауссова функция плотности вероятности (ГФПВ). Значения параметров, входящих в ГФПВ (среднее значение $\bar{x}$ и ковариационная матрица Σ), вычисляются из набора пикселей с использованием стандартных методов. В результате вероятность того, что новый пиксель $\vec{x} = (H, S)$ соответствует по цвету кожи, может быть вычислена как

$$P(\vec{x}) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x} - \bar{x}) \times \Sigma^{-1} \times (\vec{x} - \bar{x})^T \right). \quad (2)$$

В конечном итоге результатом процесса сегментации будет представление руки в виде реперной области путем применения алгоритма объединения пикселей, удовлетворяющих выражению (2). Полученные таким образом результаты являются инвариантными к фону и различным условиям освещения.

4 Механизм классификации (моделирования)

Для каждого изображения рук вычисляются его главные компоненты, согласно ранее описанному алгоритму. Обычно берётся от 5 до 200 главных компонент. Процесс распознавания заключается в сравнении главных компонент неизвестного изображения с компонентами всех остальных изображений. Для этого обычно применяют какую-либо метрику (простейший случай – Евклидово расстояние). При этом предполагается, что изображения рук сгруппированы в кластеры в собственном пространстве. Из базы данных (или тренировочного набора) выбираются изображения-кандидаты, имеющие наименьшее расстояние от входного (неизвестного) изображения.

Для классификации можно использовать метод сравнения эталонов (Template Matching), который заключается в выделении областей рук на изображении, и последующем сравнении этих областей для двух различных изображений. Каждая совпавшая область увеличивает меру сходства изображений. Для сравнения областей используются простейшие алгоритмы, вроде попиксельного сравнения. Недостаток этого метода заключается в том, что он требует много ресурсов как для хранения участков, так и для их сравнения. В виду того, что используется простейший алгоритм сравнения, изображения должны быть сняты в строго установленных условиях.

Выводы

В статье рассмотрены методы автоматического распознавания жестов рук и предложен алгоритм распознавания жестов рук для реализации удобного интерфейса управления автоматизированными системами. С учетом того, что управление происходит в режиме реального времени, предлагаются технологии, обеспечивающие быстродействие и простоту алгоритма и не требующие огромных вычислительных затрат. Основываясь на возможности дальнейшего применения в робототехнике, предлагается определенный набор жестов, который используется при обучении системы, а также формировании соответствующих команд. В качестве набора жестов рекомендуется использовать жесты из алфавита глухонемых.

Алгоритм можно модифицировать, если включить дополнительные этапы предобработки, такие как калибровка камер, фильтрация и.т.п. Этап сегментации достаточно прост и при использовании в сложных условиях работы нуждается в усовершенствовании. Надежное выполнение алгоритма распознавания жестов предполагает учет неоднозначной природы статических и динамических жестов, проблем выделения руки на изображении, условий освещения и помех.

Литература

1. Абакумов В.Г. Интерпретация движений рук расширяет возможности интерактивного управления в интеллектуальных системах / В.Г. Абакумов, Е.Ю. Ломакина // Природные и интеллектуальные ресурсы Сибири. – 2009. – С. 199-202.
2. Davis J. Visual Gesture Recognition / J. Davis and M. Shah // EE. Proc.-Vis. Image Signal Process. – April 1994. –Vol. 141, № 2.
3. Chang C.-C. Hand Pose Recognition Using Curvature Scale Space / C.-C. Chang, I.-Y. Chen and Y.S. Huang // IEEE International Conference on Pattern Recognition. – 2002.
4. Utsumi A. Multi-Camera Hand Pose Recognition System Using Skeleton Image / A. Utsumi, T. Miyasato and F. Kishino // IEEE International Workshop on Robot and Human Communication. – 1995. – P. 219-224.
5. Rosales R. 3D Hand Pose Reconstruction Using Specialized Mappings / R. Rosales, V. Athitsos, L. Sigal and S. Sclaroff // IEEE International Con. on Computer Vision. – 2001. – P. 378-385.
6. Freeman W.T. Orientation Histograms for Hand Gesture Recognition / W.T. Freeman and M. Roth // IEEE International Conf. on Automatic Face and Gesture Recognition. – 1995.
7. Bretzner L. Hand Gesture Recognition using Multi-Scale Color Features, Hierarchical Models and Particle Filtering / L. Bretzner, I. Laptev, and T. Lindberg // IEEE International Conf. on Automatic Face and Gesture Recognition. – 2002.
8. Kato M. Articulated Hand Tracking by PCA-ICA approach / M. Kato, Y.W. Chen, and G. Xu // Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition. – Southampton, 2006. – P. 329-333.
9. Manresa C. Hand tracking and gesture recognition for human-computer interaction / C. Manresa // Electronic Letters on Computer Vision and Image Analysis. – 2005. – № 5(3). – P. 96-104.

В.Г. Абакумов, Є.Ю. Ломакіна

Автоматичне розпізнавання жестів в інтелектуальних системах

У роботі розглянута концепція використання жестів рук в інтелектуальних системах, зокрема для керування робототехнічними пристроями. Проаналізовано основні процедури розпізнавання рухів рук, виділені основні моменти на кожному з етапів розпізнавання. Запропоновано підхід, що дозволяє зменшити розмірність простору, у якому перебуває рука, а також уникнути обмежень, пов'язаних з використанням спеціальних маркувальних засобів.

V.G. Abakumov, E.Yu. Lomakina

Automatic Gesture Recognition in Intelligent Systems

The article describes the concept of hand gesture application in intelligent systems to the control robotic devices is considered. The basic procedure of hand motion recognition was analyzed in detail, and the main points in each recognition stage were emphasized. An approach to reduce the space dimensionality in which hand is located and to avoid the limitations associated with the use of special markings has been proposed.

Статья поступила в редакцию 01.06.2010.